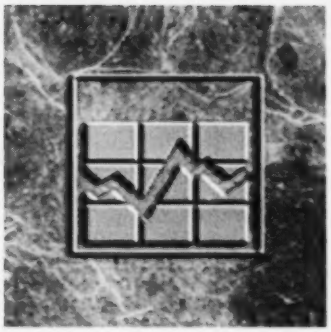
N° 75F0002MIF au catalogue — N° 003
ISSN : 1707-2867
ISBN : 978-0-662-09282-7

Document de recherche

Série de documents de recherche - Revenu

La qualité des données dans l'Enquête sur la dynamique du travail et du revenu (EDTR) en 2005



par Christopher Duddek

Division de la statistique du revenu
Immeuble Jean-Talon, Ottawa, K1A 0T6

Téléphone: 1-613-951-7355

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Toute demande de renseignements au sujet du présent produit ou au sujet de statistiques ou de services connexes doit être adressée à : Division de la statistique du revenu, Statistique Canada, Ottawa, Ontario, K1A 0T6 (téléphone : 613-951-7355; 888-297-7355: revenu@statcan.ca).

Pour obtenir des renseignements sur l'ensemble des données de Statistique Canada qui sont disponibles, veuillez composer l'un des numéros sans frais suivants. Vous pouvez également communiquer avec nous par courriel ou visiter notre site Web à www.statcan.ca.

Service national de renseignements	1-800-263-1136
Service national d'appareils de télécommunications pour les malentendants	1-800-363-7629
Renseignements concernant le Programme des services de dépôt	1-800-700-1033
Télécopieur pour le Programme des services de dépôt	1-800-889-9734
Renseignements par courriel	infostats@statcan.ca
Site Web	www.statcan.ca

Renseignements pour accéder au produit

Le produit nº 75F0002MIF au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.ca et de choisir la rubrique Publications.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées dans le site www.statcan.ca sous À propos de nous > Offrir des services aux Canadiens.

Statistique Canada
Division de la statistique du revenu

Série de documents de recherche - Revenu

La qualité des données dans l'Enquête sur la dynamique du travail et du revenu (EDTR) en 2005

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2007

Tous droits réservés. Le contenu de la présente publication électronique peut être reproduit en tout ou en partie, et par quelque moyen que ce soit, sans autre permission de Statistique Canada, sous réserve que la reproduction soit effectuée uniquement à des fins d'étude privée, de recherche, de critique, de compte rendu ou en vue d'en préparer un résumé destiné aux journaux et/ou à des fins non commerciales. Statistique Canada doit être cité comme suit : Source (ou « Adapté de », s'il y a lieu) : Statistique Canada, année de publication, nom du produit, numéro au catalogue, volume et numéro, période de référence et page(s). Autrement, il est interdit de reproduire le contenu de la présente publication, ou de l'emmagasiner dans un système d'extraction, ou de le transmettre sous quelque forme ou par quelque moyen que ce soit, reproduction électronique, mécanique, photographique, pour quelque fin que ce soit, sans l'autorisation écrite préalable des Services d'octroi de licences, Division des services à la clientèle, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Mai 2007

Nº 75F0002MIF au catalogue, vol. 3
ISSN 1707-2867
ISBN 978-0-662-09282-7

Périodicité : hors série

Ottawa

This publication is available in English upon request (catalogue no. 75F0002MIF).

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Table des matières

1. Introduction 5
2. Composition/érosion de l'échantillon 6
3. Erreurs d'échantillonnage 8
4. Erreurs de couverture 9
5. Taux de réponse 11
6. Taux de permission d'accès au dossier fiscal 15
7. Taux de couplage avec le dossier fiscal 16
8. Taux d'imputation 18
9. Arrondissement des données sur le revenu 21

1. Introduction

L'Enquête sur la dynamique du travail et du revenu (EDTR) est une enquête longitudinale qui date de 1993. Elle se veut un moyen de mesurer les variations du bien-être économique des Canadiens ainsi que les facteurs touchant ces changements. La population cible est formée de toutes les personnes vivant au Canada, sauf les personnes suivantes : les résidents du Yukon, des Territoires du Nord-Ouest et du Nunavut, les résidents des réserves, les personnes institutionnalisées et le personnel militaire en casernement.

L'échantillon de l'EDTR est formé de 2 panels. Chaque panel demeure dans l'enquête pour 6 années consécutives et un nouveau panel y entre tous les 3 ans. En janvier suivant l'année de référence, les préposés aux entrevues téléphoniques prennent contact avec les ménages de l'échantillon de l'EDTR. Ils recueillent des données démographiques pour chaque membre du ménage et des données d'enquête complètes pour chaque membre de 16 ans et plus. Ils posent des questions sur le travail (l'activité sur le marché du travail, l'expérience de travail, les épisodes de chômage et l'emploi même) le niveau de scolarité et les sources de revenu.

On a apporté un changement marquant à la collecte des données de l'EDTR au cours de l'année de référence 2004. Auparavant, on disait aux répondants à la fin de l'entrevue de janvier qu'on communiquerait à nouveau avec eux en mai pour obtenir des données sur leur revenu ainsi que sur certains types de dépenses. Depuis 2004, on pose toutes les questions au cours de la même entrevue, en janvier. Le répondant peut dès lors accorder à Statistique Canada la permission d'extraire toutes les données requises du fichier des données fiscales T1, ce qui élimine la nécessité de la deuxième partie de l'entrevue. On reportait la collecte des données sur le revenu à mai dans l'espoir que le répondant (venant de produire une déclaration de revenus) connaîtrait mieux les données requises.

Bien que conçue au départ comme une enquête longitudinale, l'EDTR a toujours conservé la capacité de produire des estimations transversales. Cet aspect transversal a pris une importance nouvelle avec l'annulation de l'Enquête sur les finances des consommateurs après l'année de référence 1997. L'EDTR est alors devenue la source première de données transversales sur le revenu du ménage et de la famille.

Toutes les personnes membres des ménages choisis pour l'EDTR au début de la première année de l'existence d'un panel font partie de l'échantillon longitudinal pour l'EDTR. À ce titre, ce sont ces particuliers qui font l'objet d'une observation longitudinale. Toute autre personne vivant dans un ménage avec une personne de l'échantillon longitudinal est appelée cohabitant. Les cohabitants vivant avec des personnes de l'échantillon longitudinal admissibles à l'échantillon transversal font aussi partie de l'échantillon transversal.

Pour plus de renseignements sur les concepts, les définitions et le plan de l'enquête, voir la publication de Statistique Canada intitulée « *Enquête sur la dynamique du travail et du*

revenu – Un aperçu de l'enquête », http://www.statcan.ca :8096/bsolc/francais/bsolc?catno=75F0011X.

Les enquêtes par sondage peuvent comporter des erreurs. Comme dans toutes ses enquêtes, Statistique Canada met un temps et un effort considérables à contrôler ces erreurs à chaque stade de l'Enquête sur la dynamique du travail et du revenu. Mais il y a quand même des erreurs. Statistique Canada a pour politique de fournir des mesures de la qualité des données pour aider ses utilisateurs à bien interpréter les données. Le présent rapport résume les mesures de qualité qui ont pour objet de décrire la qualité des données de l'EDTR.

2. Composition/érosion de l'échantillon

Bien que conçue au départ comme enquête longitudinale, l'EDTR a aussi la capacité de produire des estimations transversales. Chaque personne ne faisant pas partie de l'échantillon longitudinal qui vit avec un répondant de l'échantillon longitudinal est aussi appelée à participer à l'enquête. Cette personne est dite cohabitante. Les tableaux 2.1 et 2.2 montrent la composition de l'échantillon de l'EDTR par province et par région métropolitaine de recensement (RMR), respectivement, en nombre de répondants de l'échantillon longitudinal, de répondants de l'échantillon longitudinal qui ne sont pas admissibles à l'échantillon transversal (comme les personnes décédées, les personnes institutionnalisées et celles qui ont déménagé à l'extérieur du pays), et les cohabitants répondants.

Tableau 2.1 Composition de l'échantillon de l'EDTR par province, 2005

	Taille de l'échantillon longitudinal		Échantillon longitudinal des personnes non admissibles à l'échantillon transversal[1]		Cohabitants		Taille de l'échantillon transversal	
Province	Panel 4	Panel 5	Panel 4	Panel 5	Panel 4	Panel 5	Panel 4	Panel 5
Terre-Neuve	1 305	1 468	57	11	159	87	1 407	1 544
Île du Prince Édouard	906	879	36	9	138	64	1 008	934
Nouvelle-Écosse	2 114	1 975	101	28	325	155	2 338	2 102
N.-B.	1 793	1 893	83	20	287	144	1 997	2 017
Québec	5 989	5 927	285	76	989	431	6 693	6 282
Ontario	9 189	9 689	412	165	1 285	587	10 062	10 111
Manitoba	2 234	2 352	116	36	310	133	2 428	2 449
Saskatchewan	2 251	2 495	118	36	329	165	2 462	2 624
Alberta	2 631	3 355	97	33	496	282	3 030	3 604
C.-B.	2 933	3 088	128	49	452	189	3 257	3 228
Autre	331	248	331	248	0	0	0	0
Total	**31 676**	**33 369**	**1 764**	**711**	**4 770**	**2 237**	**34 682**	**34 895**

0 zéro absolu ou valeur arrondie à zéro

1. Comprend les personnes qui sont décédées, les personnes institutionnalisées et celles qui ont déménagé à l'extérieur du pays.

Tableau 2.2 Composition de l'échantillon de l'EDTR par RMR, année de référence 2005

Région métropolitaine de recensement	Taille de l'échantillon longitudinal		Échantillon longitudinal des personnes non admissibles à l'échantillon transversal		Cohabitants	
	Panel 4	Panel 5	Panel 4	Panel 5	Panel 4	Panel 5
Halifax	450	586	84	57	534	643
Québec	424	425	126	38	550	463
Montréal	1 220	1 190	210	118	1 430	1 307
Ottawa - Gatineau	823	875	134	58	957	931
Toronto	1 548	1 745	257	127	1 805	1 872
Hamilton	388	446	65	24	453	470
St. Catherines - Niagara	454	447	70	21	524	468
Kitchener	445	462	59	35	504	497
London	400	498	55	38	455	536
Windsor	317	376	54	16	371	392
Winnipeg	1 008	1 233	165	87	1 173	1 320
Calgary	611	721	126	52	737	773
Edmonton	598	1 016	130	88	728	1 104
Vancouver	1 004	1 071	159	51	1 163	1 122
Victoria	298	308	49	25	347	333
Other CMA or CA	10 854	11 469	1 746	805	12 600	12 270
N'habitez pas dans un RMR	9 070	9 797	1 281	597	10 351	10 394
Pas disponible[1]	1 764	704	0	0	0	0
Total	**31 676**	**33 369**	**4 770**	**2 237**	**34 682**	**34 895**

0 zéro absolu ou valeur arrondie à zéro
1. Ces renseignements ne sont disponibles que pour les personnes qui sont admissibles à l'échantillon transversal.

La couverture de l'échantillon transversal de l'EDTR est maintenue par l'ajout de cohabitants chaque année. La seule exception est le cas des immigrants qui arrivent après le début d'un panel et avant le début du suivant, et qui forment de nouveaux ménages, ce qui introduit une légère sous-représentation. L'échantillon longitudinal, toutefois, est sujet à érosion. L'érosion est la perte graduelle de répondants chaque année pendant la vie du panel. Le tableau 2.3 montre le statut de répondant pour les personnes choisies au départ dans l'échantillon longitudinal. Le tableau montre aussi que l'échantillon longitudinal des répondants est formé des répondants faisant partie du champ de l'enquête, des personnes qui ont déménagé au Yukon, dans les Territoires du Nord-Ouest ou au Nunavut, des personnes qui ont déménagé à l'extérieur du Canada, des personnes institutionnalisées et des personnes décédées.

Tableau 2.3 Statut des personnes de l'échantillon longitudinal, année de référence 2005

Statut des personnes de l'échantillon longitudinal	Panel 4	Panel 5
Dans le champ de l'enquête (répondants)	29 912	32 665
Hors du champ de l'enquête (non-répondants)	2 568	8 958
Déménagé au Yukon, dans les T.N. O., au Nunavut	5	5
Déménagé à l'extérieur du Canada	324	241
Institutionnalisées	467	186
Décédées	968	272
Éliminées de l'échantillon[1]	7 953	3
Doubles/erreurs[2]	35	0
Total	**42 232**	**42 330**

0 zéro absolu ou valeur arrondie à zéro

1. Les répondants peuvent être retirés de l'échantillon pour deux raisons. Les ménages dont tous les membres ont refusé de répondre pour 2 cycles consécutifs sont classés " refus catégoriques ", et il ne se fait pas d'autres efforts pour les dénombrer. De même, après deux ans, on ne tente plus de nouveaux essais de collecte des données auprès des ménages dont on a perdu la trace.
2. Répondants qui ont été inclus par erreur dans le ménage au début de la première année de l'existence d'un panel.

3. Erreurs d'échantillonnage

Les erreurs d'échantillonnage viennent de ce que les inférences au sujet de la population observée sont fondées sur les données d'un échantillon de cette population plutôt que de l'ensemble de la population. Le plan d'échantillonnage, la variabilité des caractéristiques mesurées et la taille de l'échantillon sont autant de facteurs qui contribuent à l'ampleur de l'erreur d'échantillonnage.

L'erreur type est une mesure fréquente de l'erreur d'échantillonnage. L'erreur type mesure le degré de la variation introduite dans les estimations par le choix d'un échantillon particulier plutôt que d'un autre de la même taille et du même plan. Une autre mesure fréquente de l'erreur d'échantillonnage est le coefficient de variation (CV), qui est l'erreur type estimée exprimée en pourcentage de l'estimation.

Dans l'EDTR, on utilise l'approche « bootstrap » pour le calcul des erreurs types. C'est une méthode de rééchantillonnage pour l'estimation de la variance, souvent appliquée aux estimations d'un plan d'échantillonnage complexe. Le tableau 3.1 fait voir les niveaux de CV provinciaux et nationaux pour un échantillon d'estimations clés de l'EDTR.

Tableau 3.1 Coefficients de variation (%) nationaux et provinciaux pour certaines variables de l'EDTR

Variable	T.N.L.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.	Canada
Revenu total moyen	1,6	1,8	1,4	1,2	1,1	1,2	1,6	1,6	1,3	1,6	0,6
Revenu moyen du marché	2,0	2,6	1,5	1,5	1,3	1,4	1,8	1,9	1,5	1,8	0,7
Salaires et traitements moyens	2,1	2,2	1,5	1,4	1,0	1,4	1,4	1,2	1,3	1,8	0,7
Prestations moyennes d'a. e.	4,3	4,9	6,7	4,4	3,6	3,8	6,5	7,5	6,8	5,9	1,8
Aide sociale moyenne	5,8	11,1	6,5	6,1	4,2	3,5	9,3	9,9	7,1	5,0	2,2
Autre revenu moyen	24,2	16,0	7,6	12,0	8,4	7,3	11,8	9,8	8,2	14,7	4,4
Familles sous le SFR (après impôt)	9,3	13,5	6,7	6,5	3,9	3,4	5,6	5,5	5,6	5,1	1,9
Nombre de personnes employées	1,6	2,2	1,0	1,4	1,1	0,9	1,2	1,1	1,2	1,3	0,5

4. Erreurs de couverture

Pour qu'une enquête produise de bonnes estimations, il faut que son échantillon soit représentatif de la population observée. Pour ce faire, on utilise les projections démographiques du recensement comme totaux de contrôle pour corriger les poids de l'EDTR. Le taux de glissement est une mesure de la différence en pourcentage entre ces projections du recensement et l'estimation de l'enquête par utilisation des poids avant l'application de ce rajustement pour le glissement. Plus spécifiquement, le glissement se calcule par la formule suivante :

$$glissement_c = \frac{\left(CP_c - \sum_{k \in S_c} w_{kc}\right)}{CP_c} * 100$$

où la classe C est la classe ou le groupe pour lequel nous voulons calculer les taux de glissement. Par exemple, au niveau détaillé, les groupes sont fondés sur la province, le sexe et le groupe d'âge.
CP_C est la projection démographique du recensement pour la classe C
w_{kc} est le poids de l'enquête pour la kième unité répondante de la classe C
S_C est l'ensemble de ménages-échantillon dans la classe C

Les taux de glissement pour les enquêtes-ménages sont généralement positifs en raison de la sous couverture de la base de sondage.

Le graphique 4.1 montre les taux de glissement historiquement. Le tableau 4.1 montre les taux de glissement au niveau des personnes par province et par panel.

Les taux de glissement ont aussi été calculés au niveau des ménages et sont résumés au tableau 4.2 et graphique 4.2.

Figure 4.1 Taux de glissement au niveau des personnes (%)

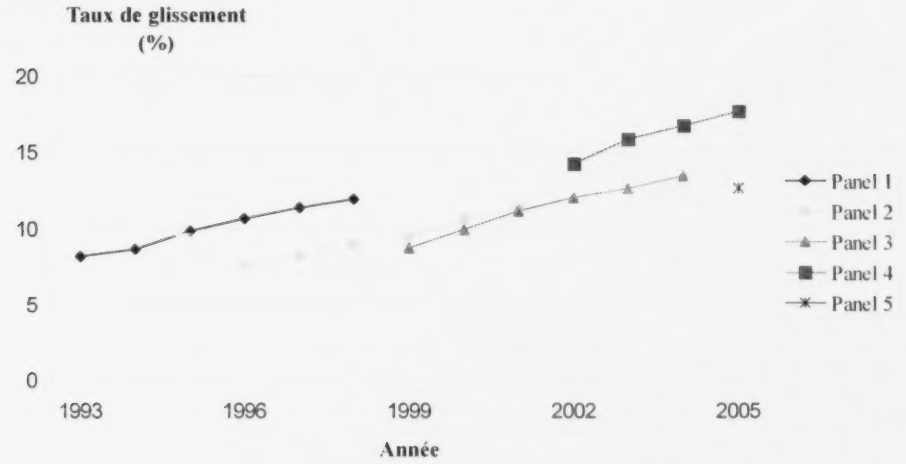


Tableau 4.1 Taux de glissement au niveau des personnes par province de 2005 (%)

	T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.	Canada
Panel 4	10,1	5,7	12,0	12,5	14,2	18,9	15,4	11,0	19,4	24,9	17,7
Panel 5	2,9	7,0	6,4	3,6	5,6	18,0	2,6	0,5	18,5	14,5	12,7

Figure 4.2 Taux de glissement au niveau des ménages (%)

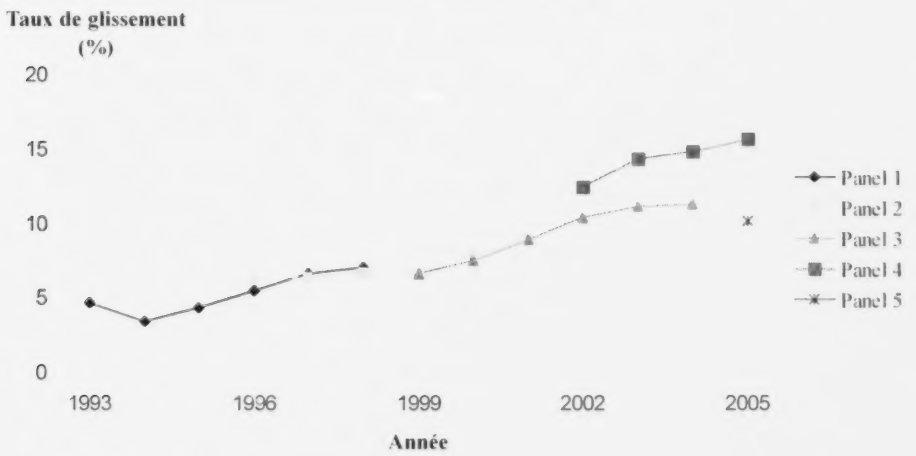


Tableau 4.2 – Taux de glissement au niveau des ménages pour les provinces par taille du ménage (%)

Province	Panel 4 Taille des ménages				Panel 5 Tailles des ménages e			
	1	2	3 ou plus	Tous	1	2	3 ou plus	Tous
Terre-Neuve	24,5	4,2	7,1	9,7	5,5	3,7	-1,9	1,7
Île-du-Prince-Édouard	8,9	-0,3	3,7	3,5	11,0	2,9	4,9	5,7
Nouvelle-Écosse	7,4	11,4	11,4	10,4	2,2	11,7	2,4	5,8
Nouveau-Brunswick	1,1	10,7	14,3	9,6	13,3	0,6	0,5	3,7
Québec	17,3	16,6	11,3	15,0	7,5	6,4	4,3	6,0
Ontario	14,2	14,6	18,4	16,2	13,4	15,8	16,3	15,4
Manitoba	15,0	3,8	20,5	13,3	-17,1	1,1	4,1	-3,2
Saskatchewan	-18,2	15,3	15,8	5,8	6,6	-12,1	3,1	-1,2
Alberta	-3,1	13,8	22,4	13,2	-7,1	21,7	18,2	13,3
Colombie-Britannique	20,5	24,8	23,8	23,2	1,4	11,1	15,4	9,9
Canada	**13,2**	**15,6**	**17,4**	**15,6**	**6,3**	**11,1**	**11,8**	**10,1**

5. Taux de réponse

Puisque l'EDTR vise à fournir des estimations longitudinales et transversales, les taux de réponse doivent être calculés pour ces deux aspects. Les taux de réponse transversaux sont calculés à la fois au niveau des personnes et au niveau des ménages. Comme les personnes faisant partie de l'échantillon ont l'option d'accorder la permission qu'on consulte leur dossier fiscal et d'éviter ainsi les questions sur le revenu, il est possible d'avoir des données complètes sur le revenu sans communiquer avec le répondant pendant l'année de référence. La définition d'un non-répondant n'est par conséquent pas une définition simple.

Si toutes les personnes d'un ménage constituent des non-répondants aux questions sur le travail et le revenu, ces personnes (et les ménages) seront classées en tant que non-répondants.

En ce qui concerne les personnes des ménages non-répondants aux questions sur le travail dans le cas desquelles nous disposons des données fiscales, on détermine si la personne fait partie du même ménage que l'année précédente (au 31 décembre). Si le ménage n'est pas le même, cela signifie que le répondant s'est séparé du ménage initial. Puisque nous n'avons pas le moindre renseignement sur la composition du nouveau ménage, ces personnes sont définies comme non-répondantes.

Les personnes des ménages non-répondants aux questions sur le travail dans le cas desquelles nous disposons toutefois des données sur le revenu et dont le ménage est le même que l'année précédente sont considérées comme des non-répondants si le ménage était un ménage non-répondant aux questions sur le travail en janvier précédent. Comme on effectue la collecte des mises à jour sur la composition des ménages en même temps que sont posées les questions sur le travail, on n'aura pas mis à jour la composition des

ménages pendant deux années consécutives. Les personnes des ménages qui n'ont pas répondu aux questions sur le travail au cours de deux collectes de janvier consécutives sont en conséquence considérées comme des non-répondants de l'EDTR.

Le graphique 5.1 illustre les taux de réponse à l'EDTR des personnes de l'échantillon transversal au fil des années de l'enquête. On calcule les taux de réponse au niveau des personnes en divisant le nombre de répondants admissibles à l'échantillon transversal aux questions sur le travail et/ou le revenu par le nombre total de personnes admissibles à l'échantillon transversal. On suppose que les non-répondants sont encore dans la population cible, à moins d'indication contraire. Par conséquent, cela peut donner une certaine sous-estimation des taux de réponse.

Le ménage est considéré comme ménage répondant si au moins une personne du ménage est considérée comme répondant. On calcule les taux de réponse des ménages en divisant le nombre de ménages répondants admissibles à l'échantillon transversal par le nombre total de ménages admissibles à l'échantillon transversal. Encore une fois, on fait une hypothèse : on suppose que les ménages non répondants sont encore dans la population cible. Cela peut donc amener une certaine sous-estimation des taux de réponse.

La non-réponse risque d'introduire un biais dans les données. Il y a un biais lorsque les caractéristiques des répondants diffèrent de celles des non-répondants et que cette différence a un effet sur la variable observée. Il est difficile d'établir si la non-réponse introduit un biais étant donné que la quantité de renseignements disponibles pour les non-répondants est limitée. Le graphique 5.2 illustre les taux de réponse des ménages de l'échantillon transversal par province.

Figure 5.1 Réponse des personnes de l'échantillon transversal (%)

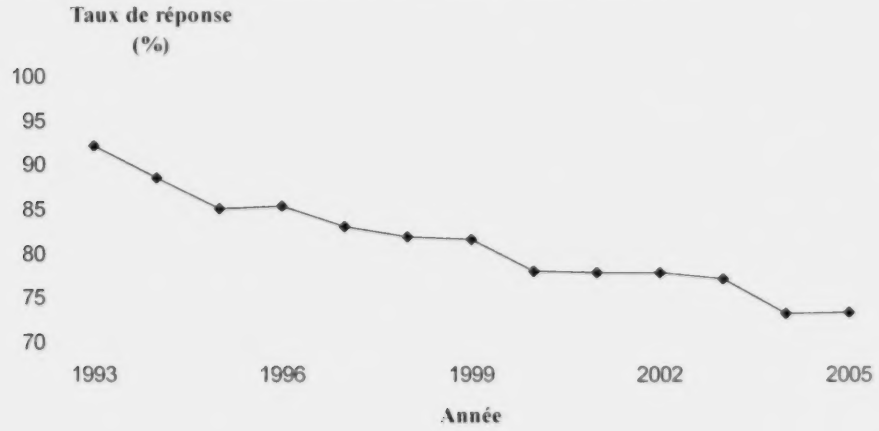
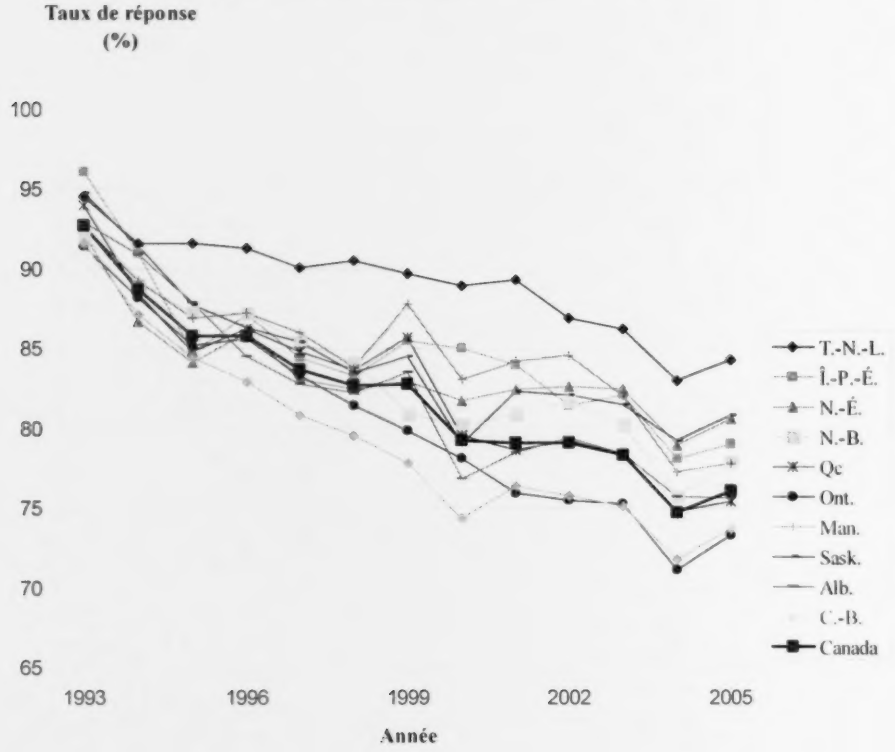


Figure 5.2 Réponse de ménages de l'échantillon transversal (%)



Le tableau 5.1 illustre les taux de réponse des personnes par phase. Les pourcentages de « répondants aux questions sur le travail » et de « répondants aux questions sur le revenu » correspondent aux pourcentages de personnes qui ont répondu seulement aux séries de questions sur le travail ou sur le revenu, respectivement, alors que les « répondants aux deux séries » représentent le pourcentage de personnes qui ont répondu en totalité ou en partie aux deux séries de questions.

Étant donné qu'il est difficile de définir conceptuellement le ménage de l'échantillon longitudinal, seuls les taux de réponse longitudinaux au niveau des personnes sont calculés. Le tableau 5.2 montre ces derniers pour chacun des quatre panels. On calcule ces taux en divisant le nombre de répondants de l'échantillon longitudinal par le nombre initial de personnes de l'échantillon longitudinal choisies dans ce panel.

Tableau 5.1 Taux de réponse des personnes de l'échantillon transversal par phase[1] (%)

Année	Répondants aux deux séries	Répondants aux questions sur le travail	Répondants aux questions sur le revenu	Non-réponse
1993	75,6	10,3	6,2	7,9
1994	75,1	10,5	2,8	11,6
1995	71,7	10,0	3,3	14,9
1996	71,6	10,8	2,9	14,6
1997	68,9	12,2	2,2	16,7
1998	68,8	10,4	2,6	18,2
1999	65,5	13,6	2,5	18,5
2000	56,1	17,3	4,6	22,0
2001	63,3	10,4	4,1	22,2
2002	61,6	10,8	5,4	22,2
2003	63,9	7,9	5,4	22,9
2004	62,3	5,8	5,1	26,8
2005	62,1	8,3	2,9	26,7

1. Taux de réponse longitudinaux par panel (personnes de tout âge).

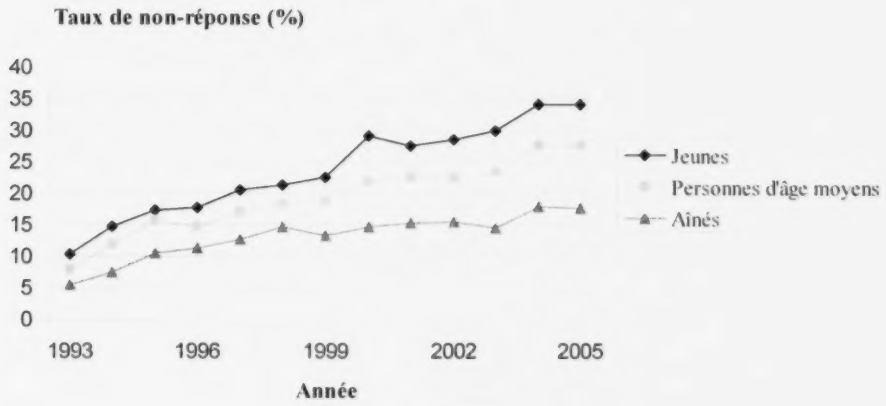
Tableau 5.2 Taux de réponse des personnes de l'échantillon longitudinal (%)

	Cycle de Panel					
Panel (date de début)	1	2	3	4	5	6
Panel 1 (1993)	93,3	89,6	86,5	83,9	82,4	81,5
Panel 2 (1996)	89,5	86,7	85,2	82,7	78,5	77,4
Panel 3 (1999)	83,9	83,0	83,0	79,6	76,4	73,7
Panel 4 (2003)	81,2	83,2	78,3	75,0	...	...
Panel 5 (2005)	78,8	...	...	...	...	...

... n'ayant pas lieu de figurer

Le graphique 5.3 illustre les taux de non-réponse longitudinale enregistrés chaque année, par groupes d'âge. Les « jeunes » sont des personnes ayant au moins 16 ans mais plus de 30 ans. Les « personnes d'âge moyen » correspondent aux personnes de 30 ans ou plus mais de moins de 60 ans, et les « aînés » sont les personnes âgées d'au moins 60 ans.

Figure 5.3 Non-réponse longitudinale par groupes d'âge



6. Taux de permission d'accès au dossier fiscal

Avant l'année de référence 2004, on tenait deux entrevues par année : l'entrevue de janvier concernait des activités comme le travail, la fréquentation scolaire, la recherche de travail ou la retraite. La seconde entrevue, en mai, touchait le revenu, mais elle n'était pas nécessaire si le répondant accordait à Statistique Canada la permission d'obtenir les données requises de son dossier fiscal. La source fiscale devrait livrer des données cohérentes de grande qualité, de sorte qu'un taux élevé de permission devrait assurer des estimations de revenu de grande qualité. Statistique Canada demandait cette permission au répondant à la fin de l'entrevue de janvier. Si le répondant refusait, on communiquait encore une fois avec lui en mai. À ce moment-là, on lui demandait de nouveau s'il préférerait donner la permission qu'on accède à son dossier fiscal. S'il refusait, l'entrevue avait lieu. Depuis l'année de référence 2004, on demande cette permission seulement une fois, en janvier. Si le répondant n'accorde pas sa permission, l'entrevue se poursuit aussitôt avec les questions sur le revenu.

Le tableau 6.1 montre les taux de permission par panel pour chaque phase de l'enquête. L'option de donner la permission de consultation du dossier fiscal a été donnée pour la première fois lors de la collecte de mai pour l'année de référence 1994. Avant cela, toutes les données sur le revenu étaient recueillies par voie d'entrevue.

Les pourcentages du tableau sont fondés sur le nombre de répondants de plus de 15 ans qui sont admissibles à l'échantillon transversal. Les taux de permission des années de référence 2000 et 2004 sont les mêmes pour janvier et mai, parce qu'il n'y a pas eu d'entrevue en mai ces années-là.

Tableau 6.1 Taux de permission pour l'EDTR par panel

	Cycle											
	1		2		3		4		5		6	
	Entrevue		Entrevue		Entrevue		Entrevue		Entrevue		Entrevue	
Panel (date de début)	Janv	Mai	Janv	Mai	Janv	Mai	Janv	Mai	Janv	Mai	Janv	Mai
Panel 1 (1993)	...	...	...	58.5	69.7	75.4	78.8	82.4	83.5	85.1	86.1	88.3
Panel 2 (1996)	53.2	68.9	76.0	79.3	83.3	86.3	85.8	88.2	87.5	...	87.0	90.3
Panel 3 (1999)	55.0	71.7	76.4	...	78.2	84.7	86.6	90.0	89.8	91.9	92.7	...
Panel 4 (2002)	58.3	77.3	80.6	85.4	88.5	...	89.9	...	...	...	...	...
Panel 5 (2005)	69.7	...	...	...	...	...	...	...	...	...	...	...

... n'ayant pas lieu de figurer

Note : La permission a été demandée pour la première fois en mai 1994.
En 2000, la collecte de mai sur le revenu a été remplacé par une deuxième collecte des données sur le travail.
En raison des coûts de collecte élevés, les entrevues sur le revenu ont été abandonnées en 2004.

7. Taux de couplage avec le dossier fiscal

Les répondants peuvent accorder à Statistique Canada la permission d'utiliser leurs données fiscales, mais on ne leur demande pas leur numéro d'assurance sociale (NAS). Sans NAS pour repérer les répondants à l'EDTR dans le fichier de données fiscales, il faut effectuer une opération de couplage pour trouver le NAS du répondant. Le Système généralisé de couplage d'enregistrements (SGCE) mis au point à Statistique Canada sert à effectuer ce couplage.

Après un traitement préliminaire du fichier de données fiscales et du fichier de l'EDTR pour vérifier la compatibilité du formatage de toutes les variables d'appariement, on effectue un appariement direct à l'aide de 7 principales variables d'appariement. Ces variables d'appariement sont : le sexe, la province, le code Soundex[1] pour le nom de famille, le nom de famille, la date de naissance, le code postal et la première initiale. L'enregistrement de l'EDTR ne peut pas avoir de données manquantes pour les variables d'appariement clés. Les résultats de l'appariement direct sont soumis à une opération manuelle de détection des erreurs lorsqu'un enregistrement de l'EDTR correspond à plus d'un enregistrement fiscal, lorsque plus d'un enregistrement fiscal correspond à un enregistrement de l'EDTR, et lorsque le premier prénom n'est pas le même dans les 2 sources (seule la première initiale est utilisée dans l'appariement). Le taux d'appariement dans l'appariement direct est d'environ 55 %.

Les dossiers non appariés sont alors soumis à appariement statistique. On définit des cases[2] pour l'appariement. Les fichiers sont segmentés en cases, où les variables

1 .Soundex est le nom d'une routine de codage du nom qui sert à éliminer les erreurs d'orthographe fréquentes dans les noms de famille des répondants. Ce codage s'effectue en fonction de la prononciation du nom de famille.

2. Les cases sont des groupes de particuliers dans le fichier de données fiscales et le fichier de l'EDTR, pour lesquels le sexe, la province et le code Soundex sont les mêmes.

déterminantes sont le sexe, la province et le nom de famille en code Soundex. Chaque enregistrement d'une case dans le fichier de l'EDTR est comparé à chaque enregistrement de la même case dans le fichier fiscal. On attribue des facteurs d'importance pour la concordance parfaite, la concordance partielle et la non-concordance.

Ces facteurs sont des valeurs numériques, qui servent à évaluer la probabilité qu'une paire d'enregistrements (un de l'EDTR et un du dossier fiscal) représente la même personne. On définit des facteurs pour chacune des variables d'appariement. On définit des seuils où l'on détermine qu'il y a concordance certaine entre les enregistrements si leur facteur total dépasse le seuil supérieur ou non-concordance certaine si leur facteur total est inférieur au seuil inférieur. On effectue une vérification manuelle pour s'assurer de la qualité des appariements.

Figure 7.1 : Pourcentage de NAS trouvés

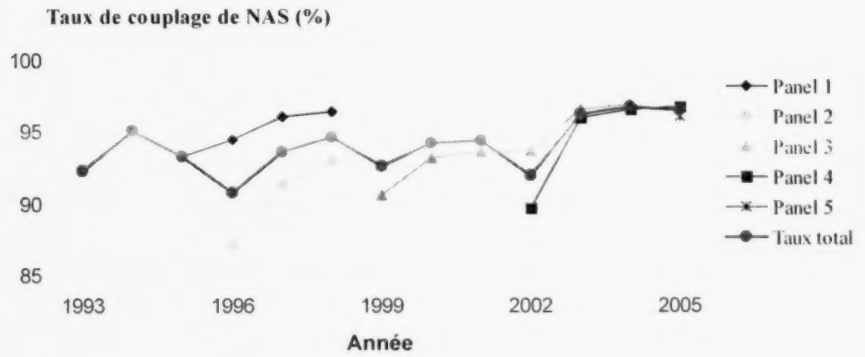
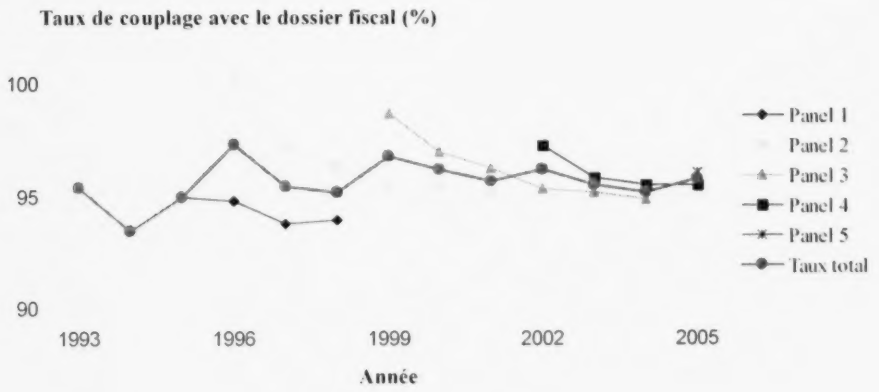


Figure 7.2 : Taux de couplage EDTR/dossier fiscal



Le graphique 7.1 donne le pourcentage des répondants de l'échantillon de l'EDTR qui donnent leur permission et pour lesquels un NAS peut être trouvé. Puisque certains répondants qui donnent leur permission n'ont pas produit de déclaration de revenus, il n'est pas possible d'obtenir un couplage positif dans tous les cas où un NAS est trouvé. Le graphique 7.2 donne les taux de couplage pour les personnes de l'échantillon de l'EDTR pour lesquelles nous avons réussi à trouver un NAS.

Le tableau 7.1 compare la proportion d'enregistrements fiscaux avec ceux recueillis au cours de l'entrevue téléphonique. Au total, dix-huit variables du revenu sont imputées pendant l'imputation du revenu dans l'EDTR. Pour un grand nombre de personnes, une imputation partielle suffit. L'imputation est dite partielle lorsqu'un ou plusieurs postes de revenu sont imputés à l'aide de renseignements fournis par le particulier.

Tableau 7.1 Proportion des répondants venant du dossier fiscal ou de l'entrevue (%)

Année	Dossier fiscal	Entrevue	Autres[1]
1999	71,9	12,0	16,2
2000	74,0	0,0	26,0
2001	78,9	5,0	16,1
2002	74,2	8,8	17,0
2003	81,4	5,2	13,4
2004	83,4	5,0	11,7
2005	73,6	9,8	16,6

1. Comprend les enregistrements qui ne sont pas couplés avec le dossier fiscal et sans réponses aux questions sur le revenu.

8. Taux d'imputation

Pour compenser pour les ménages non répondants de l'échantillon de l'EDTR, on applique une correction de la non-réponse aux poids de l'EDTR. Cependant, on conserve dans l'échantillon les ménages qui ont donné une réponse partielle et on impute les données sur le revenu qui sont manquantes pour des personnes des ménages qui ont répondu. Il peut être nécessaire de procéder à une imputation complète de toutes les variables du revenu de ces personnes, ou seulement de certaines zones. On peut considérer que les taux d'imputation dans l'EDTR sont une mesure de la réponse partielle à l'enquête.

L'imputation des variables du revenu dans l'EDTR s'effectue par la méthode du plus proche voisin. On définit un ensemble de règles de cohérence de base et, pour un enregistrement donné nécessitant l'imputation, un ensemble de donneurs. On définit également un ensemble de variables d'appariement, dont chacune est corrélée avec les variables à imputer. Par l'utilisation combinée d'une fonction de score (pour les variables d'appariement catégoriques) et une fonction de distance (pour les variables

d'appariement numériques), on trouve l'enregistrement donneur le plus semblable au receveur et on utilise pour imputer les données de l'enregistrement.

Le tableau 8.1 indique le pourcentage des personnes des ménages répondants de l'EDTR qui ont été soumises à une imputation totale ou partielle. Il faut se rappeler qu'un ménage répondant de l'EDTR est un ménage qui compte au moins un membre ayant répondu en partie ou en totalité aux questions sur le travail ou sur le revenu de l'enquête.

Tableau 8.1 Personnes pour lesquelles des variables du revenu doivent être imputées, par province en 2005 (%)

Province	Imputation totale[1]	Imputation partielle[2]	Pas d'imputation
Terre-Neuve	2,3	21,8	75,9
Île du Prince Édouard	2,2	20,0	77,8
Nouvelle-Écosse	1,6	22,4	76,0
Nouveau-Brunswick	2,1	22,0	75,9
Québec	2,3	19,6	78,1
Ontario	3,5	26,2	70,2
Manitoba	2,1	23,2	74,7
Saskatchewan	1,8	24,2	74,0
Alberta	3,1	24,3	72,6
Colombie-Britannique	3,3	24,7	72,1
Canada	**2,7**	**23,5**	**73,8**

1. Pas de renseignements fournis par le répondant. Tous les éléments de données sont imputés.
2. Un ou plusieurs éléments de données sont imputés à l'aide de certains renseignements fournis par le répondant.

Au tableau 8.2, nous comparons le pourcentage des enregistrements de données fiscales nécessitant une imputation avec le pourcentage des enregistrements pour lesquels les données sont recueillies par l'entrevue téléphonique. On détermine la nécessité d'imputer partiellement la section du revenu après avoir combiné les réponses aux questions sur le travail et sur le revenu. On corrige les cas incohérents lors du processus d'imputation. Le tableau 8.2 montre également le pourcentage des personnes soumises à l'imputation partielle pour lesquelles il faut imputer entre une et dix-sept variables.

Tableau 8.2 Enregistrements fiscaux et enregistrements d'entrevue nécessitant une imputation totale ou partielle (%)

	Source des données			
Imputation	Fiscaux	Entrevue	Autres[1]	Tous
Partielle (1 variable)	6,2	14,5	0,0	6,0
Partielle (2 à 9 variables)	0,2	34,2	0,0	3,5
Partielle (10 à 17 variables)	0,0	0,0	13,9	13,9
Totale	...	...	2,7	2,7
Pas d'imputation	93,5	50,9	...	73,8
Total	**100,0**	**100,0**	**100,0**	**100,0**

... n'ayant pas lieu de figurer

1. Enregistrements qui n'ont pu être appariés aux données fiscals sans information sur le revenue. Certains de ces cas ont partiellement été imputés à l'aide d'information tirée de l'entrevue sur le travail.

En 2002, en collaboration avec la Société canadienne d'hypothèques et de logement (SCHL), on a ajouté à l'EDTR de nouvelles questions sur le logement au profit de la recherche et de l'élaboration des politiques de logement. L'enquête réunit désormais les renseignements suivants depuis l'année de référence 2002 : la nécessité de réparations (selon l'occupant du logement); le combustible principal utilisé pour le chauffage du logement; et la présence d'une ferme ou d'une entreprise dans la propriété. Également, dans le cas des propriétaires-occupants : le montant des paiements hypothécaires périodiques; le montant des impôts fonciers annuels; et si le logement est un logement en condominium. Dans le cas des locataires, on recueille les variables suivantes : le montant du loyer mensuel, les commodités incluses dans le loyer (*p. ex.*, chauffage, eau, électricité); et si le loyer est subventionné par le gouvernement ou un employeur.

Tableau 8.3 Ménages nécessitant une imputation des variables du logement, par province en 2005 (%)

Province	Imputation totale[1]	Imputation partielle[2]	Pas d'imputation
Terre-Neuve	4,1	31,8	64,1
Î. P. É.	6,3	33,3	60,4
Nouvelle-Écosse	4,5	31,1	64,4
Nouveau-Brunswick	13,6	22,1	64,3
Québec	3,6	29,6	66,8
Ontario	2,8	39,7	57,5
Manitoba	3,3	38,4	58,2
Saskatchewan	4,2	38,1	57,7
Alberta	3,1	41,1	55,9
Colombie-Britannique	1,6	42,7	55,7
Canada	**3,9**	**36,0**	**60,2**
Canada (2004)	**4,0**	**31,3**	**64,7**

1. Pas de renseignements fournis par le répondant. Tous les éléments de données sont imputés.
2. Un ou plusieurs éléments sont imputés à l'aide de certains renseignements fournis par le répondant.

Tableau 8.4 Enregistrements nécessitant une imputation partielle (%)

	Nombres des variables de ménages qui ont besoin d'imputation			
Année	1	de 2 à 5	de 6 à 19	Un ou plus
2004	10,5	9,9	10,9	31,3
2005	10,2	10,1	15,7	36,0

Les renseignements qui précèdent s'ajoutent aux renseignements sur la propriété et le type de logement (depuis 1994) et aux renseignements sur la présence d'un prêt hypothécaire et le nombre de chambres à coucher dans le logement (depuis 1999).

À cause de la non-réponse à certaines questions, il a fallu commencer à imputer des variables du logement dans l'EDTR en 2002. Deux méthodes d'imputation ont été

utilisées : l'imputation longitudinale et l'imputation transversale par donneur. L'imputation transversale par enregistrement donneur utilise une méthode semblable à celle utilisée pour l'imputation du revenu, faisant appel à la fonction de pointage décrite plus haut. Le tableau 8.3 montre le pourcentage des ménages répondants de l'EDTR qui ont été soumis à une imputation totale ou partielle.

Au total, vingt variables du logement sont imputées dans le cadre de l'imputation des variables du logement dans l'EDTR. Un grand nombre de ménages ne nécessitent qu'une imputation partielle. Le tableau 8.4 montre la répartition de ceux qui nécessitent une imputation partielle.

9. Arrondissement des données sur le revenu

Un faible pourcentage de données sur le revenu dans l'EDTR proviennent des données recueillies dans le cadre d'une entrevue téléphonique. On peut considérer que la plupart des données obtenues du fichier de données fiscales sont cohérentes mais on dénote que la qualité des données obtenues par voie de collecte comporte certaines lacunes. Un certain nombre de répondants peuvent se reporter à leurs documents d'impôt et donner des montants précis, mais il est possible qu'un grand nombre des réponses soient des estimations ou des approximations, donc exprimées en centaines ou milliers de dollars plutôt qu'en dollars et en cents.

Pour vérifier la présence possible d'un arrondissement, on a produit des distributions de chacune des tranches des 4 derniers chiffres des variables déclarées. On s'attendrait normalement que la distribution soit à peu près uniforme, et que les chiffres 0 à 9 forment chacun environ 10 % de la distribution.

Tableau 9.1 Distribution des quatre derniers chiffres des salaires et traitements supérieur à 9 999$ en 2005 (%)

	Quatrième chiffre de la fin (%)		Troisième chiffre de la fin (%)		Avant-dernier chiffre (%)		Dernier chiffre (%)	
		Dossier		Dossier		Dossier		Dossier
Chiffre	Recueilli	fiscal	Recueilli	fiscal	Recueilli	fiscal	Recueilli	fiscal
0	31,8	11,6	88,7	11,6	96,0	12,8	97,1	14,2
1	4,8	10,9	0,6	10,0	0,2	9,7	0,2	9,4
2	9,0	10,6	1,2	10,0	0,4	10,1	0,3	10,1
3	6,1	9,8	0,4	9,9	0,3	9,7	0,4	9,1
4	5,7	10,4	1,6	9,6	0,7	9,6	0,3	9,2
5	19,5	9,6	3,7	10,0	0,6	9,8	0,4	9,7
6	6,2	9,8	0,9	9,9	0,5	9,7	0,4	9,6
7	5,0	9,4	0,6	9,4	0,4	9,6	0,4	9,4
8	7,8	9,2	1,5	9,9	0,5	9,6	0,1	9,6
9	3,9	8,8	0,8	9,7	0,3	9,4	0,3	9,7

On pourrait conclure que les montants ont été arrondis à la dizaine près si les 0 sont fréquents dans le dernier chiffre, à la centaine près s'ils sont fréquents dans l'avant-

dernier chiffre également, etc. Le tableau 9.1 indique la distribution de chacun de ces chiffres pour toutes les valeurs déclarées supérieures à 10 000 de la variable « salaires et traitements », à la fois pour les données recueillies (c.-à-d. recueillies par entrevue) et pour les données fiscales. Le tableau 9.2 indique la fréquence de zéros dans chacun des 4 derniers chiffres pour toutes les valeurs autres que zéro qui ont été déclarées pour certaines variables de l'EDTR.

Tableau 9.2 Fréquence de zéros dans les 4 derniers chiffres des données déclarées pour certaines variables (%)

Variable	Quatrième chiffre de la fin	Troisième chiffre de la fin	Avant-dernier chiffre	Dernier chiffre
Salaires et traitements	26,3	81,8	94,6	96,4
Revenu de placements	8,2	34,1	68,9	86,1
Aide sociale	8,7	35,8	71,7	83,3
Prestations d'AC	6,5	47,0	87,8	94,0
Revenu de travail autonome non agricole	36,1	82,3	95,2	97,0